Replication and Robust Results

Jim Herbsleb
School of Computer Science
Carnegie Mellon University
jdh@cs.cmu.edu
http://conway.isri.cmu.edu/~jdh/
Science Is Based on a Peculiar Logic

• Experimental method
  • Relationship => hypothesis
  • Hypothesis is true
  • Conclude relationship is true

• Affirming the consequent
  • A => B
  • B is true
  • Conclude A
Many-Layered Problem

Theory

Relationship

Hypothesis

Measures
Many-Layered Problem

Theory

Relationship

Hypothesis

Measures
Robust Results

• Results consistent as “irrelevant” things vary
Multi-site Delay

Modification Request (MR) interval

Last Modification - First Modification
All changes for 2-year period

Network Element A

## Modeling Interval

<table>
<thead>
<tr>
<th>Variable</th>
<th>Measure used in models</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR interval</td>
<td>Log of number of days, first delta to last delta</td>
</tr>
<tr>
<td>Number of people</td>
<td>Log of number of people</td>
</tr>
<tr>
<td>Diffusion</td>
<td>Log of number of modules touched by change</td>
</tr>
<tr>
<td>Size</td>
<td>Log of number of delta</td>
</tr>
<tr>
<td>Time</td>
<td>Date</td>
</tr>
<tr>
<td>Severity</td>
<td>Is high severity</td>
</tr>
<tr>
<td>Fix</td>
<td>Is fix</td>
</tr>
<tr>
<td>Multi-site</td>
<td>Set of sites of all actors has more than one element</td>
</tr>
</tbody>
</table>
H1 Multi-site work just takes longer
H2 Multi-site MRs are larger, take longer
H3 Multi-site MRs are more diffuse, take longer
H4 Multi-site MRs involve more people, take longer
Graphical model of work interval for Network Element A
Multi-site work just takes longer.

2. Multi-site MRs are larger, take longer.

3. Multi-site MRs are more diffuse, take longer.

4. Multi-site MRs involve more people, take longer.
1. Multi-site work just takes longer

2. Multi-site MRs are larger, take longer

3. Multi-site MRs are more diffuse, take longer

4. Multi-site MRs involve more people, take longer
1. Multi-site work just takes longer

2. Multi-site MRs are larger, take longer

3. Multi-site MRs are more diffuse, take longer

4. Multi-site MRs involve more people, take longer
1. Multi-site work just takes longer
2. Multi-site MRs are larger, take longer
3. Multi-site MRs are more diffuse, take longer
4. Multi-site MRs involve more people, take longer
The Decision . . .

- Published in ICSE
- What next?
  - Declare victory and move on?
  - Replicate with different data?
- What was different?
  - Locations
  - People
  - Product
  - Software type
Graphical model of work interval for Network Element A (left) and B (right)
Graphical model of work interval for Network Element A (left) and B (right)
Thoughts on Replication

- Replicating the result was a bit scary
  - What do we do if the results are different?
  - But that’s science
- How similar must results be?
Graphical model of work interval for Network Element A (left) and B (right)
Closer? More Differentiated?

• Would we have learned more from a closer replication?
• From a more differentiated replication?
  • Differentiated how?
  • What would we have learned?
Replication is Always about Generalization

- Close replication
  - Generalize over concrete instances
- Differentiated replication
  - Generalize over additional variables
- External replication
  - Generalize over experimenters/labs
What Do You Learn?

Effect unlikely to be anomalous

Original result was anomalous

Same result: Robust effect

Different result: Many possible causes

Closer

More differentiated

Amount Of Learning
Most of Science is Replication

Theory

Relationship

Hypothesis

Measures